Army Research Laboratory

Iconic Representation and Dynamic Information Fidelity: Implications for Decision Support

by Robert P. Mahan, Jun Wang, Nancy Yanchus, Linda R. Elliott, Elizabeth S. Redden, and Ruby Shattuck

ARL-CR-0580 **October 2006**

prepared by

University of Georgia
AHRP Laboratory
Department of Psychology
Athens, Georgia 30602

under contract

DAAD19-01-C-0065

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-CR-0580 **October 2006**

Iconic Representation and Dynamic Information Fidelity: Implications for Decision Support

Robert P. Mahan, Jun Wang, Nancy Yanchus, and Ruby Shattuck
University of Georgia

Linda R. Elliott and Elizabeth S. Redden
Human Research and Engineering Directorate, ARL

prepared by

University of Georgia
AHRP Laboratory
Department of Psychology
Athens, Georgia 30602

under contract

DAAD19-01-C-0065

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
October 2006

2. REPORT TYPE
Final

3. DATES COVERED (From - To)
January 2005 to January 2006

4. TITLE AND SUBTITLE
Iconic Representation and Dynamic Information Fidelity: Implications for Decision Support

5a. CONTRACT NUMBER
DAAD19-01-C-0065

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Robert P. Mahan, Jun Wang, Nancy Yanchus, and Ruby Shattuck (all of Univ of GA); Linda R. Elliott and Elizabeth S. Redden (both of ARL)

5d. PROJECT NUMBER
62716-AH70

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Georgia
AHRP Laboratory
Department of Psychology
Athens, GA

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-CR-0580

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The contracting officer's representative (COR) is Linda Elliott, U.S. Army Research Laboratory, ATTN: AMSRD-ARL-HR-MW, Bldg 459, Aberdeen Proving Ground, MD 21005-5425.

14. ABSTRACT
This report addresses the fitness of using icon-based systems to support decision making within a domain where the quality and reliability of the information are dynamically changing. Specifically, the study examined the use of iconic and non-iconic display forms to communicate information fidelity levels to decision makers for the purpose of supporting land navigation judgments. The results indicated that graphic and animated icons, as well as traditional digital display formats, produced accurate navigation judgments when information fidelity was high. In contrast, graphic and animated icon formats produced highest performance when information fidelity was moderate and/or low. These results are evaluated in the context of creating useful iconic display principles that may be applied to complex and uncertain decision environments where the fidelity of the information used to make decisions is in flux.

15. SUBJECT TERMS
decision making; icons; information display; uncertainty; visual icons

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Linda R. Elliott
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (*Include area code*)
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	45	706-542-3056

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK

1. Introduction

Of all the analogies that can be used to represent real-time human-machine interaction and control in closed loop systems, the idea of the human operator as an error-nulling device has been given the most attention (see Rouse, 1986; Vicente & Rasmussen, 1992; Rasmussen, 1988). It is often referred to as the servo-mechanism analogy and defines the cybernetic paradigm developed by Norbert Wiener (1948). Here, an operator interacts with a system and adjusts control behavior on the basis of error characteristics contained in system responses to control operations. This analogy can be applied to most any problem where an operator can change system functioning through operations performed on its parameters. In addition, it presents a unique framework for understanding a special complexity affecting awareness and control in some operational environments. For example, how does one maintain calibrated decision rules learned while performing an unreliable and uncertain task when these rules must be executed in the absence of decision feedback? That is, can we create an interface that provides information to facilitate error-nulling goals during conditions when the fidelity of system measurements is in flux, along with an absence of information pertaining to actual system outcomes?

1.1 Threats to Decision Accuracy: Feedback Loops and Uncertainty

There appears little doubt today that the quality of complex decision making is directly related to the nature of feedback loops that are used by a decision maker to alter decision strategies in order to maintain decision accuracy. Feedback about decision outcomes helps maintain properly calibrated decision rules, particularly in uncertain decision domains (Balzer, Doherty, & O'Connor, 1989; Brehmer, 1974, 1978; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). In addition, feedback reduces the "out-of-the-loop" performance problem that leads to operator failure at problem detection and control, which occurs when operators lose their ability to understand the relationship among system parameters and optimal decision behavior in automated work environments (Moray, 1986; Wickens, 1992; Edwards & Less, 1974).

However, an issue for developing decision support techniques for real-time operational decision tasks is related to the fact that many systems cannot provide timely feedback to the decision maker about the quality of their decision making. Furthermore, within most real-world judgment tasks, information about the criterion is often delayed, absent, or unusable (Hammond, 1996). For example, some decision domains are limited to simple dichotomous outcomes (e.g., correct/incorrect), which often do not assist the decision maker in understanding the relation between outcomes and decision information. Simple outcome feedback does not provide sufficient detail for the operator to understand the complex and probabilistic relationship between information sources and decision criteria. Variability in criteria as a function of particular decision information arrays and the fact that identical arrays can give rise to different criterion values makes it difficult to discover causal relations. This fact underscores why relying on outcome feedback alone to train people to perform probabilistic decision tasks can be ineffective (see Hammond, 1996, for detailed review). Finally, the absence of usable decision feedback is particularly problematic for tasks defined by (a) numerous information sources, (b) dynamic environments, (c) time pressure constraints on executing judgment policies, and (d) tasks associated with high levels of fatigue and stress (for a review of unique factors related to complexity in decision making, see Endsley & Kiris, 1995; Klein, Orasanu, Calderwood, & Zsambok, 1993; Mahan, 1994; Dunwoody, Marino, Mahan, & Haarbauer, 2000; Marino & Mahan, in press; Orasanu & Salas, 1993; Parasuraman, Molloy, & Singh, 1993; Wickens, 1992; Zsambok & Klein, 1997).

1.2 Probabilistic Character of Operational Systems

In many natural decision environments, decision makers are faced with the task of assimilating information sources of limited or changing fidelity. In the present case, reliability refers to the consistency of indicator source variables used to measure system states (Woods, 1988; Rasmussen, 1988; Stewart, 2001). These indicators can refer to the instruments, observations, algorithms, automation, and actual display systems that are used to measure and represent system information (Stewart & Lusk, 1994; Parasuraman, Molloy, & Singh, 1993; Vicente & Rasmussen, 1992; Wickens, Gempler, & Morphew, 2000). For example, the control and management of unmanned aerial vehicles (UAVs) can be enhanced by automation (Dixon & Wickens, 2004a), but they can be degraded when the automated control is unreliable (Dixon & Wickens, 2004b). Further, in tactical undersea systems, the water's salinity, depth, and temperature can systematically influence the reliability of hydrophones and other sensing devices that are used to support judgments of target acquisition and prosecution (Kirschenbaum & Arruda, 1994). In tele-robotic systems, the reliability of remotely located sensors can have serious effects on supervisory control (Sheridan, 1976; Massimino & Sheridan, 1994; Sheridan, 1992; Wiener, 1988). Similarly, in many command and control operations, it is the reliability of information that often poses the most difficulty for command-level decision making (U.S. Army Training and Doctrine Command, 1989). This is particularly true in team-based decision making where a team leader is responsible for assessing the reliability and validity of judgments made by subordinate experts who provide judgments of system criteria for the team leader to process. Here, effective team functioning is thought to be associated with the team leader's ability to manifest dyadic sensitivity or the ability to remove bias from subordinate judgments when a team-based decision is being made (Hollenbeck, Ilgen, Sego, & Hedlund, 1995; Williams & Mahan, in press). Knowledge of changes in the reliability of system measurements (whether they originate from sensing equipment, human experts, or automated decision support systems) may assist in the diagnostic use of available information, particularly in the absence of timely and usable decision feedback about the quality of decisions.

1.3 Feed-Forward Support

The concept of feed-forward support refers to the technique of providing useful decision information in advance of decision actions. In contrast to feedback, which provides information about the consequences of a decision action, feed forward informs the decision maker about information properties that theoretically can be used to aid the decision process. For example, the concept of applying cognitive feedback to train decision makers to perform probabilistic judgment tasks is largely based on a feed-forward mechanism that is geared to the probabilistic nature of the task system (see Hammond, McClelland, & Mumpower, 1981; Balzer, Doherty, & O'Connor, 1989; Brehmer & Joyce, 1988). Cognitive feedback is used to help explicate a decision maker's organizing principle when s/he is processing information. The underlying goal is to increase one's awareness and control over the properties of an implicit judgment policy by making the policy's features more explicit to the decision maker (see Hammond, McClelland, & Mumpower, 1981; Hammond, 1987; Hammond, 1996). Awareness and control over implicit decision rules used by decision makers is often enhanced by having decision makers describe and communicate how they plan to make their decisions. This form of support is typically delivered via verbal, numerical, and graphical summaries of task properties, task goals, and the decision maker's judgment policies in order to improve future decision making (Brehmer & Joyce, 1988).

1.4 Icons as Feed-Forward Information Mechanisms

The concept of encoding information in the form of an iconic representations means that theoretically, one can minimize the effort or workload necessary to assimilate the information and yet simultaneously increase the number of information channels or sources that can be processed by the operator. The former goal is achieved when perceptual processing mechanisms are employed, while the latter is achieved through careful system value mappings to multi-dimensional iconic forms. The concept of an icon display is efficient in its simplicity within the practical limits, such as portability, operational requirements, amount and tempo of information flow, and the finite cognitive resources of the user. Further, icons can be engineered to support different cognitive mechanisms that are needed for different decision tasks. Here, modifying icons in order to induce specific types of cognitive organizing principles means that designers can efficiently create representations that are congruent with the dynamically changing properties of a task or decision environment.

Experimental work is needed to ascertain how iconic instantiations may facilitate or obstruct the performance of probabilistic decision tasks. Most icon studies, although illuminating, tend to rely on subjective assessments of preference. Yet, the problem with preference studies is the absence of cumulative data describing common principles that can support icon design. Although preference data are useful, performance data are necessary to identify generalizable principles.

1.5 Visual Icons

Within the visual modality, icons can be engineered to form objects that can encode multiple information dimensions that can be parsed perceptually instead of analytically. The results of numerous studies support the use of object-like displays for enhancing a user's ability to assimilate complex information (Carswell & Wickens, 1987; Coury, Boulette, & Smith, 1989; Wickens & Andre, 1990). Some of this interest has focused on using the configural properties of object displays to support perceptual operations on analog information. In configural displays, a mapping is created so that the elemental properties of an object combine to produce an emergent feature that is representative of the integration of the elemental components. Perceptual processing appears to be most useful in information integration tasks when objects configure to produce salient emergent features (Garner, 1981; Pomerantz, 1981; Pomerantz & Pristach, 1989). Wickens and Carswell (1995) identify numerous approaches for manipulating information codes that enhance the salience of emergent features resulting from the display of multi-dimensional information arrays. For example, they show how various proximity manipulations such as code homogeneity, spatial proximity, and attribute similarity can increase the salience of emergent features, thus facilitating the perceptual processing of information. Perceptual processing is distinguished from analytical processing in the sense that information integration is more intuitive and recognition based than an intellectual exercise that requires more deliberation.

In addition to being a rapid form of processing, perceptual operations require much less effort than analytical decomposition (Wickens & Carswell, 1995). Reliance on perceptual processing tends to generate parallel-based intuitive (or holistic) forms of cognition, which, although less precise than analysis, are very robust and easy to apply (see Garner, 1974; Hammond, Hamm, Grassia, & Pearson, 1987; Hammond, 1996; Simon, 1990; Anderson, 1991; Tversky & Kahneman, 1983). Intuitive cognition tends to match well with the demands of many naturally occurring judgment tasks (Cannon-Bowers, Salas & Pruitt, 1996; Hammond, 1993, 1996).

1.6 Real-Time Decision Protocols

Trade-offs often exist between robust approximating strategies and those decision strategies geared toward analysis (see Simon, 1978, 1990; Hammond, 1993, 1996). These trade-offs are typically associated with the resources available to the user at the moment a decision is required and the immediate demands of the decision task. For example, some tasks require precise and meticulous analyses of information and are not typically suited for real-time human information processing. Analysis supports the goal of precision but at a cost of fragility; one small error renders the process imprecise. However, other tasks require the application of rapid and robust decision strategies that are less susceptible to failure. Here, importance is placed on a rapid and robust process where precision is viewed as being less critical to decision outcomes.

The question in the present study is whether the perceptual properties of iconic formats can assist an operator during situations of varying levels of uncertainty. That is, can we use perceptual organizing principles to facilitate diagnostic assessments of decision information before the act of decision making occurs? Specifically, we propose that iconic representation of information reliability will be particularly helpful when information fidelity is low or unknown. For example, medium to low information fidelity conditions should favor the representation of feed-forward information in iconic forms because diagnosticity will be a joint product of the information itself (i.e., magnitude) and information reliability. Further, the cognitive response to the configural representation should be rapid because the iconic representation induces an intuitive response in the decision maker. In contrast, highly reliable information sources may be best served with feed-forward formats that separate reliability from magnitude. In high reliability conditions, reliability and magnitude are less dependent than during medium and low reliability conditions. Thus, separating information about reliability from information about magnitude may promote analytical decomposition and induce users to produce high precision calculated judgments. Analysis would thus provide greater precision in judgment performance than inducing a deliberate computational response in the decision maker that requires more time to execute.

2. Method

The task in this study required the integration of four relatively independent information sources. Each source had two variable information elements: cue magnitude and reliability. The focus of this study was to examine the effects of cue-level iconic manipulations of reliability on multi-cue judgment performance.

2.1 Participants

Thirty-five student participants were paid volunteers for this study; 65% were female. Participants ranged in age from 22 to 26 years with a mean age of 24.3 years. None had knowledge of the experiment before the briefing that they received from the experimenter. The participants were paid $75.00 for their participation and received course credit.

2.2 Design

Several estimated regression parameters, as well as a response rate measure (time/unit judgment) were used as indices of performance and are discussed in detail next. Three levels (high-R, medium-R, and low-R) of the within-subject independent variable information reliability were crossed with four levels (numeric, graphic, animated, and no information) of the within-subjects independent variable iconic reliability presentation format. A factorial 3 x 4 repeated measures analysis of variance (ANOVA) design was used as the analytical framework with the regression values and rate measure as dependent variables. The pure repeated measures design was selected because it offered an approach for mitigating the large error terms found in many judgment studies (see Brehmer & Joyce, 1988).

2.3 Apparatus and Measures

2.3.1 Data Acquisition Environment

A Dell[1] computer was used for stimulus presentation and data collection. All training and experimental sessions took place in the Applied Psychology Simulation Laboratory at the Department of Psychology, University of Georgia. The laboratory environment was free of all time cues in order to minimize a possible response distortion because of participant anticipation of the rest breaks that were given during the study.

2.3.2 Judgment Task Simulation

The judgment task required participants to integrate information from four sources in making a set of estimates on the time in minutes required to navigate a dismounted Army platoon over terrain to a linking point with other platoons. Task sources were identified with Army land navigation scenarios (see U.S. Army Training and Doctrine Command, 1989) and included terrain, the need for stealth, concealment, and visibility. Participants formed their judgments in three information reliability conditions and across four iconic reliability feed-forward presentation formats. Cues for the judgment task were randomly generated from a hypothetical infantry land navigation task. Randomly generating the cues simplified the judgment task by producing orthogonal information dimensions and was germane to the developmental nature of the research. The selection of cue sources was based on the representation of relatively distinct variables affecting land navigation.

2.4 Fidelity Manipulation

We altered the fidelity of the navigation task partly by changing the reliability of navigation cues used to represent true navigation values. Because cue reliability is necessary in order to demonstrate fidelity in the representation of internal system parameters, the fidelity of the information acquisition process defines a relation between the objective system values available for measurement and the actual indicator values that are presented to an operator for assessment. Differences between objective and displayed system values reflect differences in the fidelity in which information acquisition occurs. Finally, unreliability in the information acquisition process has been shown to impair the quality of operator judgments of system states (Stewart, 2001; Cooksey, 1996; Wickens, Gempler, & Morphew, 2000).

2.4.1 Task Criterion

A task was constructed that produced a true value (Y) for the criterion variable navigation time that was expressed in “minutes to link up”. Task elements were taken from reports of actual

[1]Dell is a trademark of Dell, Inc.

military navigation training tasks (U.S. Army Training and Doctrine Command, 1989). Pilot experimentation was necessary to configure a task system that participants were able to master. The mathematical function defining the criterion task system (criterion model) that was selected through a series of pilot studies is described in the following equation:

$$Y' = 100 + 10(.65(\mathbf{r_1}))\, X_1 + 10(.44(\mathbf{r_2}))\, X_2 + 10(.25(\mathbf{r_3}))\, X_3 + 10(-.55(\mathbf{r_4}))\, X_4 \qquad (1)$$

in which **r** is the cue reliability coefficient and X_1 is the terrain value; X_2 is the stealth value; X_3 is the concealment value; and X_4 was the visibility cue value. In computing the true values for the criterion variable, we used a set of fixed ecological weights which are shown in equation 1. Finally, the constant values of 100 and 10 in equation 1 simply ensured that an adequate range in Y values would be generated.

2.4.2 Altering Information Fidelity

Differential cue diagnosticity was a function of the product of reliability values, fixed ecological weights, and cue magnitudes. We changed cue reliabilities by <u>randomly</u> selecting **r**-values in a specific interval from a uniform probability distribution. Altering the distribution interval changed the range of the reliabilities and thus, the average diagnostic value of a given cue source. For example, randomly selecting r-values from the interval {0.7 to 1.0} would produce cues with higher average validities than selecting r-values from the interval {0.3 to 1.0}. The reliability <u>factor</u> in this experiment thus reflected three reliability configurations that were generated with the two distribution intervals for the reliability (**r**) parameter. The aim of this factor was to determine the psychological impact of variance in the statistical reliability of cue information in relation to the judgment performance metrics.

The method selected for varying the reliabilities of individual cues produced distinct changes in the fidelity and overall predictability of navigation time when criterion values and cues were subjected to linear regression.

In the high reliability condition (high-R), all cues had (r) values sampled from the 0.7-to-1.0 interval and were on average equally reliable. When the true criterion values from the high-R task model were regressed on the cue values from a set of 40 trials, the squared multiple correlation between cue values and true criterion values was approximately 0.88 and the cues accounted for about 88% of the variance in the criterion values. This squared multiple correlation was taken to represent environmental predictability (i.e., maximum task validity).

In the medium reliability condition (med-R), the terrain cue was less diagnostic on average than in high-R because of its (r) value being sampled from the larger reliability interval (0.3 to 1.0). The other cue reliabilities remained the same as in the high-R condition. Thus, in the med-R condition, terrain was not as dependable a cue. In this case, task validity was approximately 0.77.

Within the low reliability condition (low-R), both terrain and stealth cues had the larger reliability variance interval found in the 0.3-to-1.0 interval. The remaining cues retained the original high-R reliabilities (0.7 to 1.0) interval. Task validity in low-R was approximately 0.62. Figure 2 illustrates the representation of the task, showing ecological validities (re) along with the reliability coefficients (rtd) between true cue values (T1,2,3,4) and displayed values (D1,2,3,4).

The manipulation of reliability altered the criterion model through the product of cue validity and reliability. The criteria reflected the influence of the average reliability value sampled from reliability distributions. Therefore,

$$Y' = 100 + 10(.65(\mathbf{r_1}))\ X_1 + 10(.44(\mathbf{r_2}))\ X_2 + 10(.25(\mathbf{r_3}))\ X_3 + 10(-.55(\mathbf{r_4}))\ X_4 \qquad (2)$$

became

$$\text{HighR-}Y' = 100 + 10(.55)\ X_1 + 10(.37)\ X_2 + 10(.21)\ X_3 + 10(-.47)\ X_4 \qquad (3)$$

$$\text{MedR-}Y' = 100 + 10(\mathbf{.42})\ X_1 + 10(.37)\ X_2 + 10(.21)\ X_3 + 10(-.47)\ X_4 \qquad (4)$$

$$\text{LowR-}Y' = 100 + 10(\mathbf{.42})\ X_1 + 10(\mathbf{.29})\ X_2 + 10(.21)\ X_3 + 10(-.47)\ X_4 \qquad (5)$$

The criteria reflect reliability modifications of the particular criterion terms.

2.5 Iconic Display Protocol

2.5.1 Displaying Cue Magnitude

Several distinct two-dimensional iconic geometric forms were selected for cue magnitude representation, which provided a reasonable discrimination among cues (Bailey, 1982, 1989). The iconic cue values were scaled from 1 to 10, where 1 was a small magnitude value and 10 was a large value and then mapped to the judgment interface display as follows: terrain complexity was displayed as a solid black square, stealth level was displayed as a solid black triangle, concealment level was displayed as a solid black ellipse with a horizontal major axis, and visibility was displayed as a solid black circle. The icon image forms were paired with cue constructs in an arbitrary manner. The geometric area of the cue images communicated the magnitude of the cues. The terrain, stealth, and concealment cues were all positively and independently correlated with the criterion “navigation link-up time” where 1 = very low (simple terrain, low stealth activity, low concealment activity), produced short duration link-up times, while 10 = very high (complex terrain, high stealth, high concealment) produced long duration link-up times. The visibility cue was independently and necessarily inversely related to the criterion where 1 = very low visibility produced long duration link-up times, and 10 = very high visibility produced shorter duration times. The scaling used ensured that cues presented to subjects corresponded to realistic magnitude values that one might encounter in an actual navigation task (see Gentner & Stevens, 1983, about discussions of veridical representation).

2.5.2 Displaying Reliability Information

Cue reliability information was described to the participants as representing the amount of noise in the data. The choice was made to present this information as a noise concept because pilot research showed that participants appeared to understand the notion of noise better than statistical reliability or statistical error. Thus, the information displays incorporated the complement of reliability (i.e., unreliability), which was presented as degrees of "noise" values. Here, the participants were told that noise reduced the diagnostic value of cue information. Further, the larger the noise values, the less reliable the cue, and thus the less diagnostic the cue was of navigation link-up time. Finally, the participants were told that the object of the task was to discount noisy information in judgments while increasing the weight of information that was not noisy.

Feed-forward reliability of the cues was presented to the participants in four ways:

- (1) Numeric. The unreliability (noise) value was converted into a complementary numeric percentage of a random r-value from a particular reliability condition (e.g., high-R, med-R, low-R) and displayed below each graphically presented cue magnitude value where it indicated its noise level. For example, in a given judgment trial within the high-R condition, (a) four random r-values were selected from their respective reliability intervals (e.g., 0.7 to 1.0), (b) the complements of those four values were taken (i.e., 1-r values), and (c) these values were multiplied by 100 and expressed as a percentage noise score. Thus, in the high-R condition where all reliability values were randomly selected from the 0.7-to-1.0 interval, it was the percentage complement to reliability that ranged from 30% to 0% noise that was displayed under each graphically presented cue magnitude.

- (2) Graphic. The graphically displayed cues were superimposed over a gray background image of corresponding cue geometry. In this case, noise values were mapped to the judgment display as a difference in areas between an outer image (noise) and an inner cue image (magnitude). When a cue was perfectly reliable, there was no background image (i.e., the difference in areas = 0). The larger the outer image in relation to the inner image, the greater the noise associated with the cue.

- (3) Animated Icon. Cues in this format pulsed at a frequency of 3 Hz. The amplitude of the pulse (i.e., the difference between two image area values presented as an animation) indexed the amount of noise in the cue. Here, the larger the pulse amplitude, the less reliable the cue. Thus, the animated display presented the unreliability information compared to a graphically packaged animation envelope. Here, percentage noise values were mapped to the interface as a difference between two cue magnitude images (i.e., image 1 and image 2) that were animated. For example, a 30% noise value was represented by the addition of 30% area to the image 2, making it 30% larger than image 1. When it was animated, the sensation of pulsing was seen. When a cue was perfectly reliable, it did not pulse (there was only one image).

- (4) No Reliability Information. Cue magnitude was presented to the subjects in numeric or graphic format in this condition; no information about reliability was given to the subjects. Here, the feed-forward information about reliability was absent from the judgment interface.

Figure 1 illustrates an example of the judgment interface used for the graphic condition showing the cue magnitude (black features) superimposed over a noise graphic (gray features), the difference of which informs participants of the diagnostic value of the cue. Here, the aggregate reliability computation for this set of cue features reflects a medium-R reliability level.

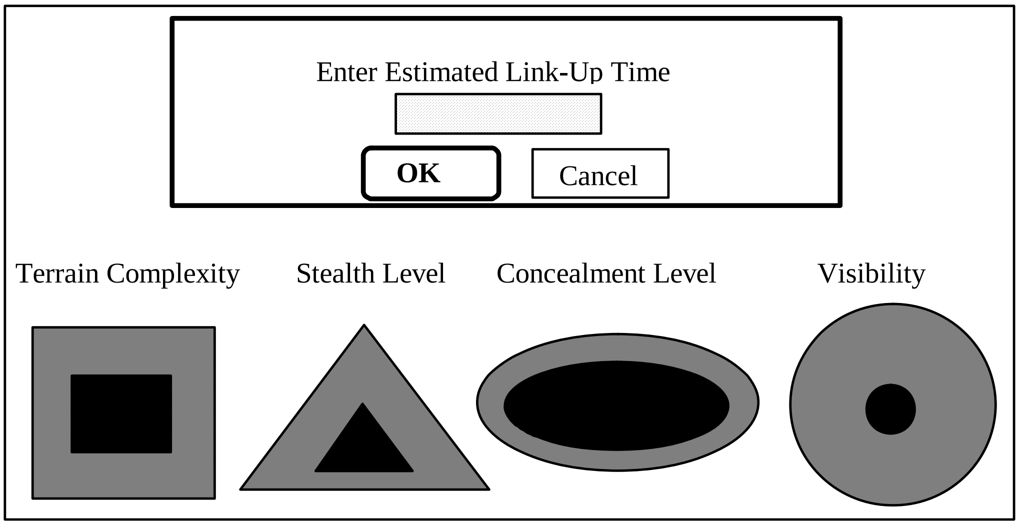

Figure 1. Static-graphic judgment interface.

2.6 Procedure

2.6.1 Training

Each participant underwent two days of intensive navigation task training, beginning at 8:00 in the morning. Each day of training consisted of a 5-hour block of time during which participants learned the judgment task. After each hour of training, participants were given a 15-minute break.

We conducted training for the navigation task by providing immediate feedback about the accuracy (outcome) of each judgment of navigation time within each reliability level (i.e., high-R, med-R, low-R). Outcome feedback consisted of the true navigation time value generated from the criterion model (navigation rule). We used cognitive feedback during training by encouraging the participants to discuss their judgment strategies with experimenters during the training process (Balzer et al., 1989). This had the effect of assisting the participants with developing an accurate organizing policy for producing criterion judgments. The cognitive feedback approach was also used to provide information to participants about the reliability of the cue information. In this case, experimenters discussed with each participant the idea of how noise might affect their ability to produce accurate judgments. In this way, participants were able to develop a conceptual sense for how cue reliability affected the judgment process, even though no other systematic reliability feedback was used during training other than what was presented with the judgment display interface.

During this training phase, each participant made navigation time estimates in the three information reliability conditions and across the information reliability presentation formats. During training, participants were told when reliability conditions were being changed. However, participants were never given specific quantitative information about the parameters of the criterion model. Participants were required to discover these parameters through trial-and-error judgments using outcome and cognitive feedback to help guide the manner in which they used the cue information. Participants had to modify their judgment protocols to incorporate changes in the reliability of navigation information. Thus, the judgment task required participants to develop strategies for weighting and integrating the cue information during different task reliability circumstances. The trial-and-error approach used here in teaching participants to discover how to make complex judgments simulates the manner in which many real-world complex judgment tasks are learned (for review, see Brehmer & Joyce, 1988).

2.6.2 Training Criterion

The training criterion was a Pearson correlation between true criterion values and judgments of navigation time equal to 90% of overall task validity. Task validity was computed as the squared multiple correlation between cues and criterion values. Participants were required to achieve 90% accuracy of the uppermost predictability of the true criterion in three consecutive 40-case trials. All 35 participants were able to meet this criterion within the 10 training hours allotted over the course of two training days. The training criterion ensured that participants had developed a set of organizing rules (policies) for judging navigation time during the various judgment conditions, which was similar, in the statistical sense, to the criterion model producing the true navigation values.

2.6.3 Experiment

All experimental sessions were conducted at the same time of day that training was administered. The experiment began the day after each participant was fully trained. Participants were presented with warm-up judgments and outcome feedback in an effort to help get them back on task. During the experiment, cue magnitude and cue noise were the only information sources available to the participants for making judgments. In order to simulate a true navigation task where the quality of judgments is not immediately known, outcome and cognitive feedback were no longer available.

During an experimental session, a single trained participant performed a subject-paced block of 40 judgments during each cue reliability and reliability presentation condition combination. Each participant performed all 12 conditions during each experimental session. After 6 of the 12 conditions were completed, each participant was given a 15-minute break. The remaining six conditions were then performed. The experiment took an average of 2 hours 15 minutes to complete. The participants received the conditions in a counterbalanced format in order to control for order effects. A statistical test for main effects attributable to the position of experimental conditions was not significant (thus, no order effects were seen), nor was a gender effect observed in the data.

3. Results

A lens model analysis of participant judgments was conducted. The lens model provides a methodological framework that incorporates the probabilistic structure of human decision ecologies and thus provides the means for modeling the relationship between a judge and criterion in a multi-cue task environment (for review, see Brehmer & Joyce, 1988; Cooksey, 1996; Hammond, 1996). Figure 2 shows the elements of the lens model, which can be mathematically characterized if the relationship among the components of the model and judgment task performance is defined. Tucker (1964) described it as follows:

$$ra = G*Rs*Re + C[(1 - Rs2)\ (1 - Re2)].5 \qquad (6)$$

The correlational performance that an individual achieves (i.e., achievement index) ra, is a function of four distinct components: the linear multiple correlation between the cue values and the criterion, Re, (environmental predictability); the linear multiple correlation between the cue values and an individual’s judgments of the criterion, Rs, (consistency index); the extent to which the linear model of the individual judge correlates with the linear model of the criterion, G, (matching index); and the extent to which the residual variance in the model of the individual correlates with the residual variance in the model of the criterion, designated C. Residual variance was negligible in this study, so the C index was not included in the analysis. A lens model representation of the judgment task in the current study is shown in figure 2.

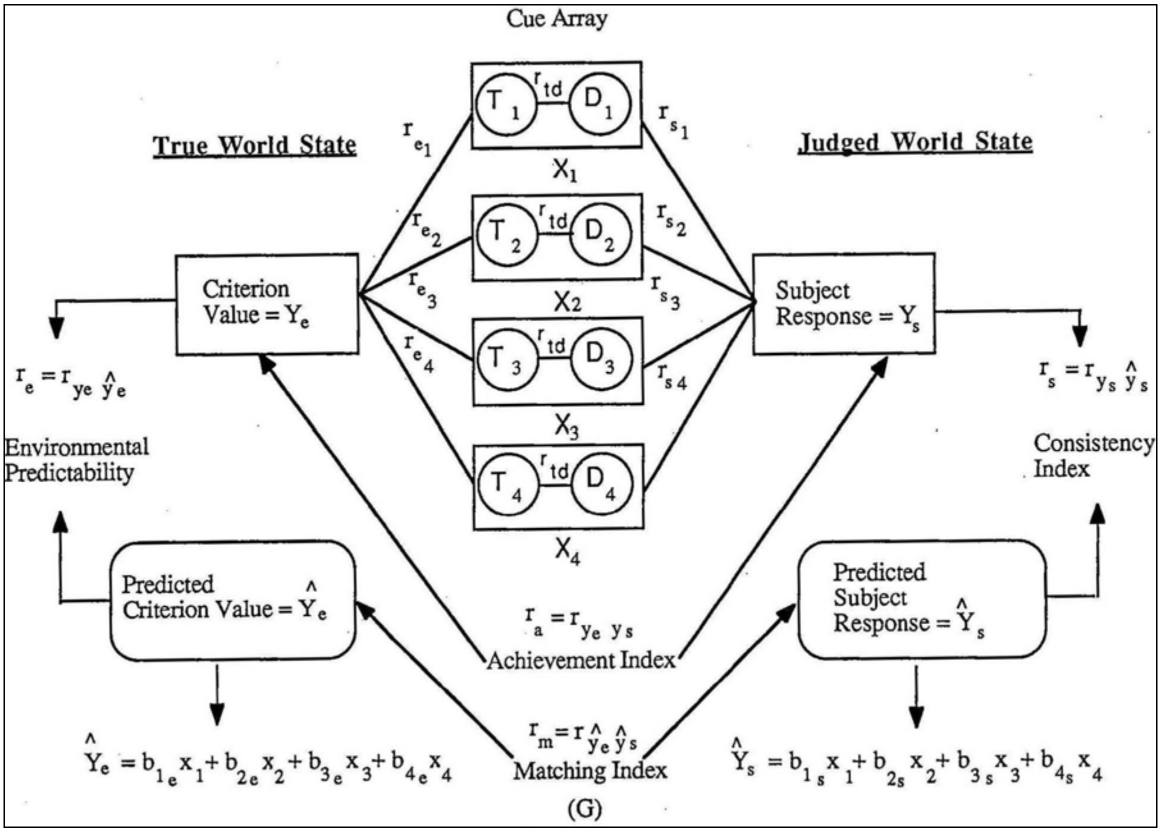

Figure 2. Lens model showing (a) the cue validities (r_{e1-4}) between cues T_1-T_4 and criterion values (true world state), (b) cue use coefficients (r_{s1-4}) between cues D_1-D_4 and judgment values (judged world state), and (c) the reliability weights (r_{td}) defining the fidelity of judged cues for representing criterion cues.

3.1 Lens Model Analysis

Graphic summaries of mean lens model indices for each of the reliability conditions across each reliability presentation format are shown in figures 3 through 5. Univariate 3 x 4 repeated measures factorial ANOVAs with three levels of cue reliability crossed with four levels of reliability presentation format were used as the analytical framework for the lens model performance indices.

3.1.1 Achievement

Achievement index scores, r_a, underwent Fisher z transformation and were then back transformed to Pearson correlations. Mean achievement index scores indicated (a) significant cue reliability main effect F (2, 68) = 12.03, $p < 0.001$; (b) significant reliability presentation main effect F (3, 102) = 14.07, $p < 0.001$; and (c) significant interaction effect F (6, 204) = 6.18, $p < 0.001$.

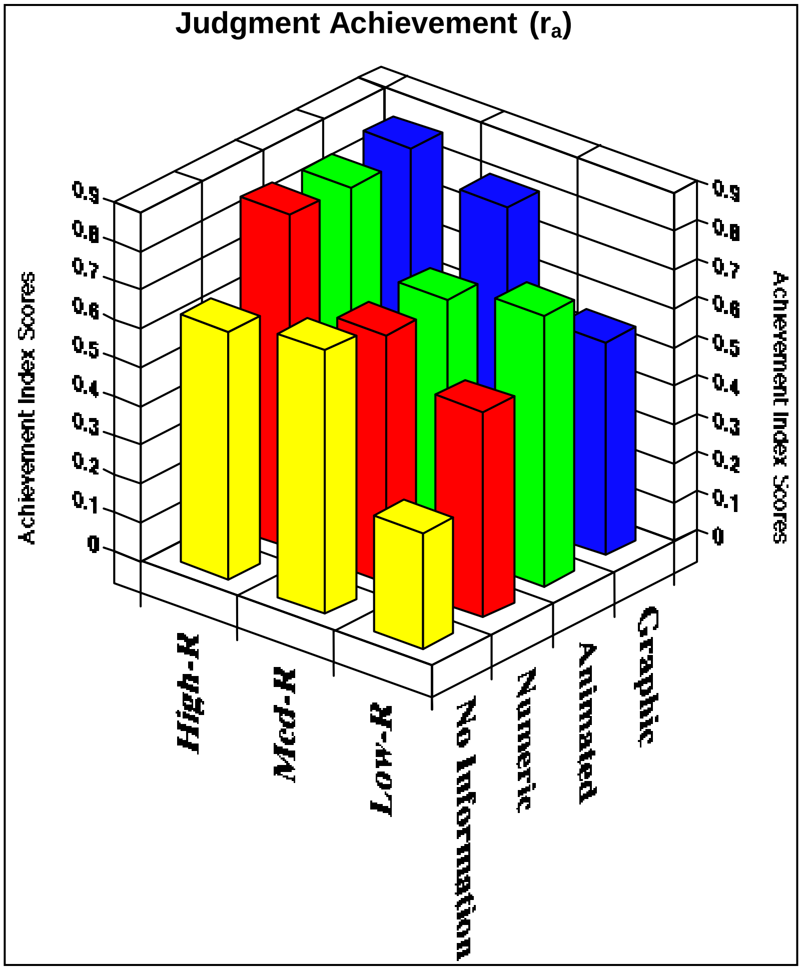

Figure 3. Achievement across display and reliability conditions.

3.1.2 Consistency

Mean consistency (R_S) scores indicated that there was (a) significant cue reliability main effect F (2, 68) = 14.11, $p < 0.001$, (b) significant reliability presentation main effect F (3, 102) = 3.89, $p < 0.01$ and (c) significant interaction effect F (6, 204) = 5.13, $p < 0.001$.

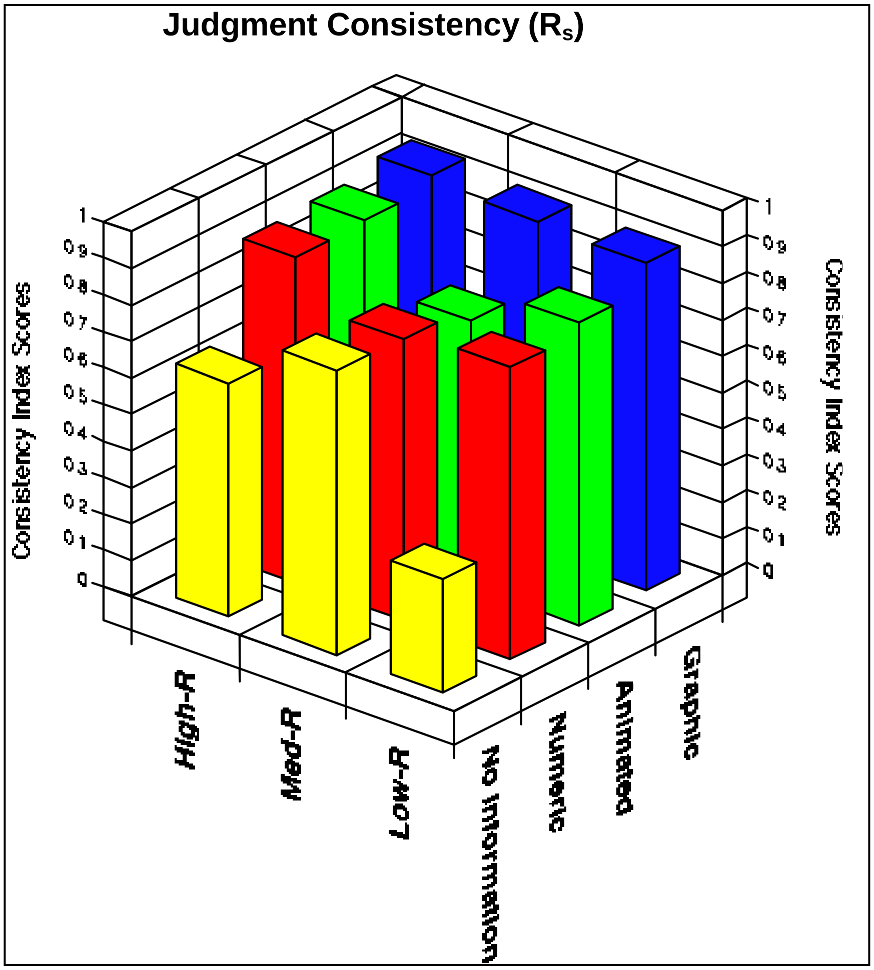

Figure 4. Consistency across display and reliability conditions.

3.1.3 Matching

Mean matching (G) scores indicate that there was a (a) significant cue reliability main effect $\underline{F}$ (2, 68) = 6.33, $p < 0.001$, (b) significant reliability presentation main effect $\underline{F}$ (3, 102) = 9.50, $p < 0.001$ and (c) significant interaction effect $\underline{F}$ (6, 204) = 2.66, $p < 0.05$.

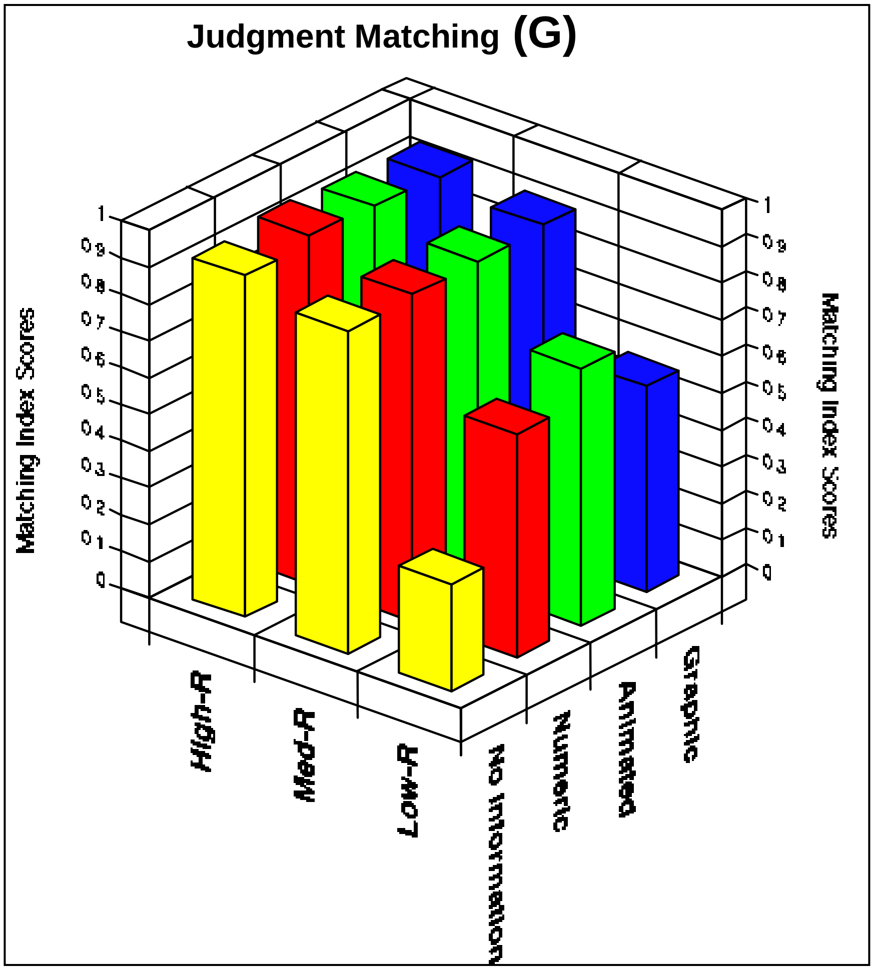

Figure 5. Matching across display and reliability conditions.

3.1.4 Judgment Response Time

The rate of judgment performance was used as a behavioral index of cognitive activity occurring in response to different experimental manipulations. The mean number of minutes necessary to complete the 40 judgment cases per participant for each condition was computed. The results of the 3 x 4 ANOVA indicate that statistically significant main effects for Reliability $\underline{F}$ (2, 68) = 4.39, $p < 0.001$ and for Display format $\underline{F}$ (3, 102) = 15.29, $p < 0.001$ were seen. Statistically significant interaction was also observed for response time $\underline{F}$ (6, 204) = 2.15, $p < 0.05$.

3.2 Parsing Factorial Interactions

3.2.1 Achievement and Consistency

Bonferroni-corrected ($p < .05$) single-degree-of-freedom (DOF) contrasts were performed on the four display formats at each level of reliability compared to a simple main-effects analysis and are displayed in table 1. Significant differences for achievement (r_a) and consistency (R_s) were found across reliability presentation formats. Pair-wise tests for the high-R condition were significant and graphic, animated, and numeric reliability presentation formats were associated with the highest achievement and consistency scores, while the no-information format produced the lowest achievement (see figure 3 for an illustration). Differences were also found among the med-R conditions, with the graphic reliability presentation format associated with the highest achievement and consistency scores, and the other formats showing no statistical differences in achievement. Finally, differences were found in the low-R conditions. In this case, the animated reliability presentation format was associated with the highest achievement scores, while the animated, graphic, and numeric scores all produced the highest consistency values.

Table 1. Bonferroni-adjusted (p <.05) simple effects comparisons of display format during high-R, medium-R, and low-R task reliability conditions.

Cognition	High R	Medium R	Low R
Achievement (r_a)	G, A, N > No*	G > A, N, No	A > G, N > No
Consistency (R_s)	G, A, N > No	G > A, N, No	A, G, N > No
Matching (G)	------------**	--------------	G, A, N > No
Response time	--------------	A > G, N, No	N > A, G, No

*G: graphic format; A: animated; N: numeric; No: no feed-forward information

**-------------- No statistical differences between formats

3.2.2 Matching

The simple main effect analysis identified the low-R reliability condition as the source of interaction for matching (G) seen in table 1. Single-DOF tests indicated that the graphic, animated, and numeric displays had higher matching values than the no-feed-forward display format.

3.2.3 Judgment Response Time

The simple main effect analysis of mean judgment response times showed that significant differences for display formats were seen within and across the reliability manipulation and are presented in table 1. Figure 6 provides an illustration of these differences. Although no differences were found in response times during the high-R condition, the animated format had the longest response time during med-R, and the numeric format produced the longest response time during the low-R condition.

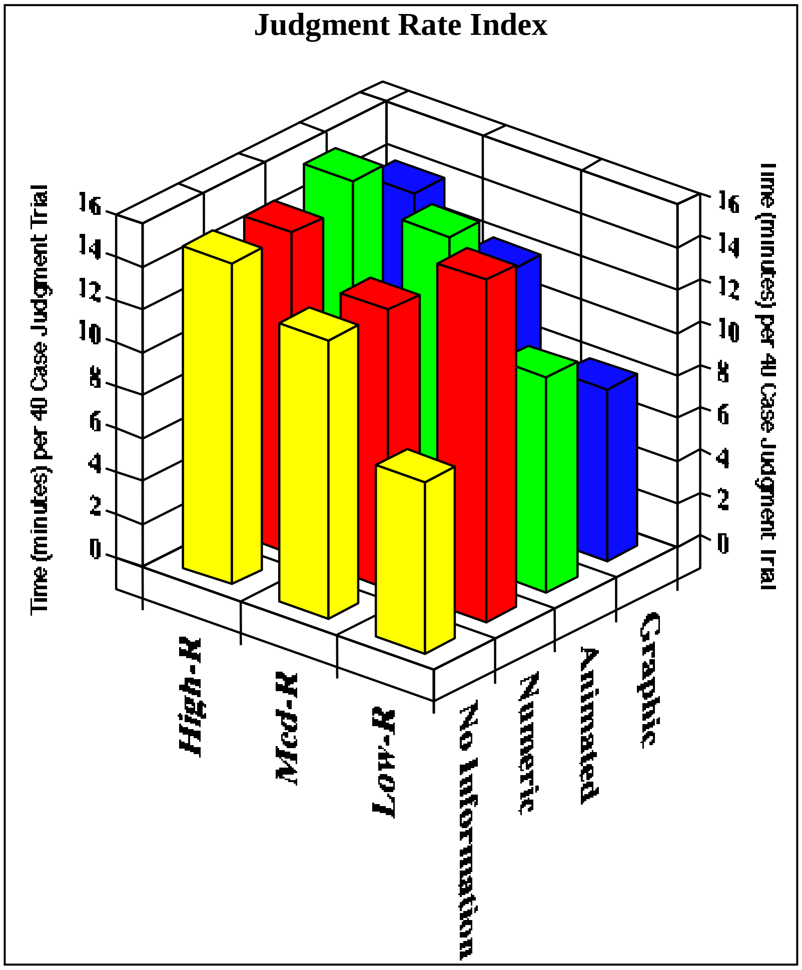

Figure 6. Judgment rate across display and reliability conditions.

3.3 Relative Importance of Cues

A descriptive assessment of the importance participants placed on various cues in generating their judgment was performed for each reliability condition across the four presentation formats. We generated bar graphs by averaging the decision policies of the participants within a reliability condition and across presentation formats. This aggregation process produced an *averaged* decision rule, which simply reflected a set of mean regression weights for each reliability condition and presentation format. Figure 7 indicates that during high-R conditions, participants saw the terrain cue as being the most diagnostic of navigation time with the concealment cue as least diagnostic across the displays presenting feed-forward information. This is consistent in its correspondence with the regression weights derived when true navigation time scores were regressed on the cue values. Further, cue use by participants was monotonic with respect to the cue validities in the criterion model for all feed-forward representations. During performance within the high-R condition, participants correctly rank ordered the cues in terms of their importance for predicting criterion scores, regardless of the display format. Thus, cue use by participants matched the linear model of the criterion.

Cue use duirng med-R appeared to display a greater average deviation from the criterion model than cue use during high-R. Figure 7 shows that cue use during med-R remained relatively monotonic across the first three reliability presentation conditions, with the terrain cue perceived as most diagnostic and the concealment and visibility cues as very similar in diagnosticity. However, a large change in cue usage is exhibited in the no-reliability information presentation condition. Finally, figure 7 displays a rather dramatic cue inversion during low-R. The least important cue (concealment) in statistically maximizing successful judgments of navigation time was weighted most, on average, by the participants across all four reliability presentation formats.

4. Discussion

Achievement performance of the judgment task displayed effects attributable to the information reliability of the task environment and the representational format in which cue unreliability or noise was communicated to the participants. In general, as the cues became less reliable, navigation time estimates became less accurate. This finding replicates many studies examining the effects of cue reliability on judgment (Brehmer, 1970; Doherty & Sullivan, 1989; York, Doherty, & Kamouri, 1987). Reliability feed-forward information, on average, facilitated judgment task performance when compared to the no-reliability feed-forward condition. The exception to this finding was the performance during low-R when participants inaccurately gave the concealment cue a large weight, treating it as a highly diagnostic information source when it actually was not. Rather than minimizing the concealment cue in link-up judgments, the participants, on average, maximized the cue in their judgments.

The graphic, animated, and numeric feed-forward reliability formats appeared equal in communicating the reliability of cues during high-R, while the graphic format appeared superior to all others in the med-R condition. Finally, the animated iconic format appeared to support judgment achievement through higher consistency scores better than other formats during low-R. Finally, the graphic iconic format appeared to produce the highest matching index scores during the low-R performance condition.

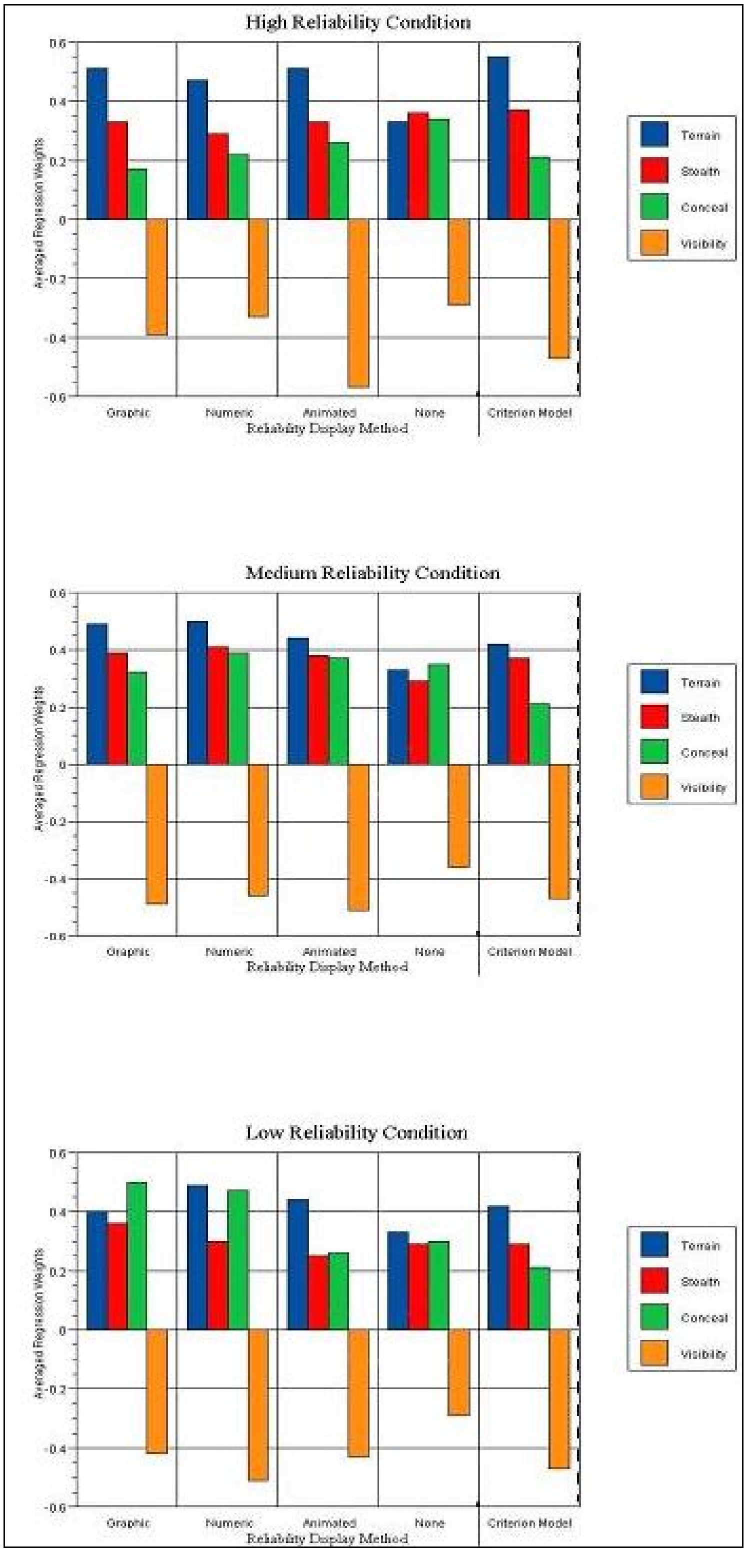

Figure 7. Use pattern duirng high, medium, and low reliability conditions and across reliability information presentation formats with the criterion model generated from "true navigation" values regressed on displayed cue values.

4.1 Cognitive Control in Task Execution

Equation 7 shows that judgment achievement (r_a) is a function of the participant's consistency (R_S), the environmental predictability (R_e), matching (G), and configural (nonlinear) residual indices of the lens model. Since the present study failed to yield a significant residual variance index (C) (by creating a strictly linear criterion model with no nonlinear components and noting an absence of nonlinear behavior in the judgment model side), the lens model equation simplified from

$$r_a = G*R_S*R_e + C[(1 - R_S^2)\ (1 - R_e^2)]^5 \qquad (7)$$

to

$$r_a = G*R_S*R\ . \qquad (8)$$

Judgment consistency refers to the reliability of a particular decision maker when s/he is executing similar decisions. That is, the consistency of his or her decision rules (e.g., cue weights) is measured by the predictability of judgments from the policy generated through least squares regression of judgments on cues. The difference between observed judgments and judgments predicted by the perfect application of one's decision policy represents decrements in consistency. Judgment consistency decrements have been viewed in the past as impairment in the ability to control the execution of a judgment policy, and thus this measure is often referred to as the control of knowledge index (Hammond, 1996; Cooksey; 1996; Hammond & Summers, 1972; Hammond & Wascoe, 1980). Hammond and Summers (1972) have argued that controlling knowledge execution can be conceptually and statistically differentiated from a decision maker's overall task knowledge state. From a conceptual viewpoint, one may intellectually understand the necessary requirements for task performance, yet be unable to actively control the application of that knowledge in task performance. These authors have used examples of executing complex motor skills in an effort to illustrate the conceptual difference between task knowledge and the implementation of that knowledge. For example, an individual may have a very good feel for what it takes (intellectually) to shoot a "free throw" in basketball (e.g., positioning oneself correctly, determining distance to basket, determining optimal ball trajectory to basket, determining force applied to the shot, understanding the importance in manifesting fluid upper body motion and fluid finishing stroke, etc.). However, being able to successfully integrate and execute these task dimensions may be quite difficult. Thus, recognizing and understanding the parameters of knowledge acquisition is not enough to guarantee successful knowledge application. One must also demonstrate cognitive control over the execution of task specific information. In addition, apart from the issue of skill in execution, there can be inconsistency in policy execution in all types of cognitive decision making. Individuals and organizations are often accused of biased decision making, that is, inconsistency in executing a particular policy across situations.

The data in the present study indicate that changes in information reliability affected the ability of judges to control the execution of their expertise in making navigation link-up time estimates. With the exception of certain aspects of the low-R condition, lower reliability translated into lower cognitive control. Reductions in cognitive control have been a general finding of judgment researchers, particularly in highly probabilistic settings (Hammond, 1981, 1996). However, lower cognitive control does not necessarily mean that participants have forgotten how to use the information—only that their ability to control execution of their judgment protocols has been altered.

4.2 Task Knowledge State

The judgment-matching index has been viewed as reflecting the participant's understanding of the properties underlying accurate task performance from the acquisition of knowledge (Brehmer & Joyce, 1988; Hammond, 1981; Hammond, Rohrbaugh, Mumpower & Adelman, 1977; Hammond & Summers, 1972). It is defined as the correlation between the linear model of the environment (the predicted criterion) and the linear model of the judge (the predicted judgment) (G). The matching index measures the extent to which participants can distinguish among the cues on the basis of their diagnostic value in predicting the criterion variable. The substantive and statistical distinction between cognitive control and task knowledge is best understood if we note that one can apply a highly consistent judgment protocol that is not empirically valid. For example, one may demonstrate a perfect application of a decision policy when no deviation exists between observed and predicted judgments. However, in this case, overall task accuracy (i.e., achievement) would be low because one is using the wrong policy for criterion judgments.

Figure 5 shows that the level of matching between participants' use of judgment information and the linear characteristics of the empirical criterion model (G) remained relatively high across the feed-forward formats during high-R and med-R conditions, and table 1 indicates the absence of statistical differences among the display formats. Within the lens model context, a high matching index means that a participant's knowledge of the task matches with the actual task system (i.e., the model derived from the regression of the true criterion values on the cues). Since training ensured that participants began the experiment knowing how to perform the task during each reliability-feed-forward condition combination, it becomes difficult to argue on the basis of the matching index for high-R and med-R that reductions in judgment achievement (r_a) seen in this study were merely a function of participants forgetting how to perform the task. For example, the loss of task knowledge often manifests the application of guessing strategies, which produce very low matching index scores. Because the matching indices remained high during these conditions, the lack of consistency in the application of decision policies was primarily responsible for reductions in achievement. There are numerous examples of situations in which an individual exhibits a high degree of understanding for the properties of a task and yet is unable to consistently apply the knowledge necessary to perform the task (see Hammond & Summers, 1972).

During performance in the low-R group, there were dramatic changes in judgment behavior. Here, the achievement (r_a) decrement was <u>not</u> associated with cognitive control measured by the consistency (R_s) index but a decrement in task knowledge that was measured by the matching index (G). Figure 4 illustrates that the consistency index remained relatively high for the reliability feedforward groups in low-R, while figure 5 illustrates the large decrement in matching (G) values for low-R performance.

Examining the cue use profiles in figure 7 provides some insight to this outcome. Although the concealment cue was predictive, on average, of navigation time in the criterion model it was viewed as highly diagnostic by the participants during low-R. Their behavior in using the concealment cue as though it were most diagnostic of navigation time led to less successful judgments. This effect is puzzling because during low-R performance, participants' cue use profiles tended to correctly discount unreliable information in their judgments. Thus, participants appeared to be aware of changes in the reliabilities compared to the feed-forward display for these cues and were able to incorporate that knowledge in their judgment policies. Here, concealment was more reliable but less important than other cues, and the participants were unable to discern this fact during low-R performance. It is very interesting and operationally relevant that participants would naturally place more importance on certain information even if it is not as important as the unreliable information.

This finding is difficult to dismiss as an experimental artifact because of its pervasive nature across the low-R feed-forward presentation formats. It is possible that in the low-R condition, the contrast between those cues presented as unreliable (i.e., large background graphics, large pulse envelope, etc.) may have made the concealment cue <u>appear</u> more valid then it was. Why the participants tended to focus on the concealment cue as a robust diagnostic source of information is difficult to explain.

However, this outcome may also reflect, in part, the general finding that people rarely achieve the level of performance found in statistical integration models (Kahneman, Slovic, & Tversky, 1982; Slovic, Fischoff, & Lichtenstein, 1977). In order to reduce the mental demands of task performance, people often execute heuristics and other resource conservation strategies for processing complex information instead of using optimizing strategies. It may simply have been easiest for the participants to use the information provided by the concealment cue and not have to encode or process reliability at all.

4.3 Iconic Feed-Forward Display and Cognition

During high-R performance, the graphic, animated, and numeric feed-forward formats demonstrated high achievement (r_a) scores and are illustrated in figure 3. Further, figure 7 demonstrates that the cue use profiles matched the weights in the criterion model during high-R. In contrast, judgment achievement during the med-R condition appeared highest for the graphic format, while the animated display appeared superior during low-R performance.

When the task was very predictable (high-R), any format could be used effectively and easily. Although it was clear in the analysis and illustrated in figure 3 that in comparison to the no-feed-forward information group, even small noise information values were useful to participants. However, the manner in which they were displayed did not matter.

There are a number of explanations for these findings that emerge from judgment research. The high reliability of the cues created what might be argued as a well-defined task, which may have led participants to apply an analytical organizing principle during performance during high-R (Hammond et al., 1987). Since task reliability was high in the high-R condition, participants could focus primarily on cue magnitude. Thus, feed-forward information allowed subjects to quickly identify that the cues were reliable and they could primarily attend to cue magnitudes. During high-R performance, participants demonstrated a capacity to execute precise judgment protocols as seen in the achievement and consistency index scores for this condition.

Some insight into the findings can be derived, in part, from feedback studies of multi-cue judgment performance. There has been substantial research suggesting that loss of control in the execution of multi-cue knowledge can be attenuated by feedback (Balzer et al., 1989; Brehmer, 1970; Doherty & Sullivan, 1989; York et al., 1987). Further, the amount and nature of feedback needed to maintain control become less demanding as the task becomes more analytical (see Hammond, 1990; Brehmer, 1978; Searcy, 1994). When the rules governing cue usage are fairly explicit (e.g., high reliability conditions), very little feedback is needed in order to stay on track or maintain control in the execution of knowledge.

However, as the task becomes more implicit because of higher levels of uncertainty produced by the addition of noise to the criterion model, feedback requirements become more demanding in order to be effective. This is, in part, why outcome feedback alone is usually insufficient in promoting task learning in multi-cue probability learning experiments that incorporate moderate to high amounts of task uncertainty (Balzer et al., 1989; Brehmer & Joyce, 1988). When tasks are well defined and certain, feedback is less important in order to promote or maintain task performance. Similarly, when information is reliable, representational properties of the feed-forward information formats are less important, and less support is necessary.

The significant main effect for reliability condition on judgment rate indicates that reliability, at least in part, affected the effort necessary to execute judgments. Figure 6 shows that on average, significantly more time was needed to complete the 40 case trials during the high-R condition than the other reliability conditions. The increased response times are a consistent feature in the application of deliberate serial processing of decision information (Kahneman & Tversky, 1982) and the application of a mental calculus to cues (Mahan, 1991; 1992; 1994; Hammond, 1990, 1996).

However, as the task's reliability factor changed, the participants began to selectively respond to the joint dependence of reliability and magnitude. During med-R, the participants seemed to use a more intuitive approach at organizing the information. They could not simply decompose the task into computing link-up times from cue magnitude information alone but were required to adopt a more general and perhaps holistic principle for aggregating the magnitude and reliability information.

During the med-R condition, reliability information had a greater impact on the diagnostic weights assigned to cues in the criterion model than during high-R. As a result, during med-R, essentially two features of the task had to be closely followed: cue magnitude and cue noise. Since cue diagnosticity was a function of both magnitude and noise, the graphic depiction may have configured the information in a manner that produced a representation that could be perceptually measured. This perceptual measurement would tend to induce an intuitive mode of organization and would best match the med-R reliability structure of the task. Hammond (1980) and others (see Garner, 1974; Hammond et al., 1987; Kubovy & Pomerantz, 1981) have noted that the reliance on perceptually measured information sources induces intuitive or holistic responses to information dimensions by people.

The reduction in response times during the med-R condition in comparison to those response times in the high-R condition seems to provide some evidence for an intuitive mode of information organization. The implication here is that the graphic feed-forward display produced the most useful mapping of task features, which called for intuitive-based judgments because of the increased noise in the task. In this case, presenting both cue magnitude and noise as superimposed images required participants to attend to, extract, and factor cue and noise values using perceptual measurement, which tends to be rapid and approximate in nature. While in the high-R condition participants could dismiss reliability as an important feature of the task, in med-R, the task required factoring both task components. Perceptual processing tends to be parallel in nature, and parallel processing has been viewed as a hallmark feature of holistic and intuitive cognitive activity (compare Hammond, 1980; 1996; Kahneman, Slovic & Tversky, 1982; Kahneman & Tversky, 1979; Kahneman & Tversky, 1982; Simon, 1978; von Winterfeldt & Edwards, 1986).

The primary decrement observed during med-R was one of exhibiting cognitive control over the execution of judgment policies. The reduction in control (measured by the consistency index R_s) that occurs in response to task performance during uncertain conditions has been observed in past research (Brehmer & Joyce, 1988; Hamm, 1988; Hammond, 1996; Hammond et al., 1987). The loss in cognitive control is believed to be a manifestation of intuitive cognition. The absence of an explicit organizing principle yields judgment protocols that randomly drift around parameter values of some optimized (normative) policy for integrating information (see Hammond, 1996, for review). However, this random drift does not necessarily compromise the overall accuracy of judgments. Within real decision environments, most information sources are significantly correlated, which of course means that the departure of a decision maker's policy cue weights from the ecological weights in a normative (criterion) model has far less impact on judgments. That is, the rank order diagnostic value of cue usage by human judges is often identical to the rank ordering of cues in the criterion model (i.e., high matching index scores), even though judgment-to-judgment variability exists in the weights applied to cues (i.e., drift). During these conditions, correlations among judgments and true values from the criterion model are typically high.

The matching data depicted in figure 5 clearly show that the knowledge of the judgment task was quite high across all feed-forward conditions for med-R performance. Once again, it was not the case of participants forgetting the diagnostic weight to be given the cues for predicting the criterion but a decline in the ability to consistently weight (i.e., factor) magnitude and reliability and integrate the diagnostic information from all cues in generating an overall judgment of link-up times.

Animated and numeric formats failed to support judgment achievement during med-R performance at the level observed in the graphic format. The explanations for these findings may be linked, at least in part, to the failure of these displays to accurately map the task. The uncertain (noisy) quality of the med-R task called for participants to respond to the joint reliability-magnitude elements compared to an intuitive approach to judgments of the criterion. However, this congruent cognitive activity did not appear to be supported in the representational character of the animated and numeric displays.

During med-R performance, the numeric display may have communicated a sense of precision to participants by 1) presenting the noise as a precise numeric quantity, and 2) separating this information from the manner in which cue magnitude was displayed. The impact of both these display features may have induced a form of analysis and the decomposition of the information sources into orthogonal parts. However, the criterion model called for configural (conditional) processing of reliability and magnitude of cues. Thus, numeric information may have induced a mode of cognition that was incongruent with the properties of the task in terms of the capacity for users to generate diagnostic assessments of the reliability and magnitude component of the cues. This observation is partially supported by the longer response times evident for the numeric display versus the graphic display in the med-R condition (see figure 6). The longer response times suggest that an analytical decomposition of the information occurred during judgment. Hammond (1980) has noted that an analysis-inducing feature of a task is the use of objective (numeric) quantities for cue values and that objective information tends to produce an analytical response in decision makers.

The animated display may have suffered similar consequences as the numeric display but for different reasons. A primary feature of the display was communicating reliability through animation. This may have required participants to analyze the size of the pulse envelope. Although the display required perceptual measurement, the differences in pulse envelopes among cues required some form of analysis for encoding. A related interpretation may be associated with the notion of salience. Within the med-R condition, the terrain cue had significant dynamic changes occurring in reliability over those reliabilities of the other cues, which meant that animation was much more visible for the terrain cue. Animation in the med-R condition may have generated a high level of salience for the unreliable cue leading to a selective focused attention aimed at encoding the animated information. This selective attention generated through the level of cue salience may have overcome any intuitive inducing features of the graphic components of the animated display, producing a shift in cognitive mode toward analysis. Once again, the longer response time data for the animated display in med-R seems to provide some evidence for the execution of an analytical strategy. The graphic display on the other hand, was more successful in med-R conditions, presumably because it more closely mapped the configural properties of the task and communicated these properties to the participants. The longer response times for the numeric and animated displays during med-R are consistent with a more analytically oriented organizing principle.

The average response characteristics of participants changed when they performed the task during conditions of low reliability (low-R). Figure 4 indicates that during low-R performance, response consistency remained relatively high over the graphic, animated, and numeric feed-forward conditions. In contrast, the low-R matching index scores appeared much more variable during this condition over the display formats (see figure 5). When the cues became least reliable, the participants' judgments, although fairly consistent, became fairly wrong. Here, judgment consistency was high, but validity was low. Thus, overall achievement (r_a) was significantly lower for the low-R condition because of the low matching (G) values in the judgment protocols (see figure 5).

During low-R performance, the animated display seemed to be the most effective of the formats in supporting the judgment process in terms of overall judgment achievement. The relatively higher achievement values for the animated display were largely attributable to the fact that during the animated condition, participants were able to generate the highest matching index values (G) of any low-R display format (see figure 5). Why the animated display was more useful to the participants during low-R when it seemed to offer poorer support during the med-R is difficult to understand. In some sense, one might expect that low levels of reliability would favor a more spatial/temporal display in order to take advantage of perceptual measurement. The success of the spatial representation approach was seen in med-R when the graphic reliability display provided the superior support for judgment. Yet, when one examines the response data, it seems that participants appeared to use the animated display in a manner that suggests perceptual encoding leading to more of an intuitive principle applied to the decision task. The finding that during med-R the animated display induced a more deliberate analysis and during low-R a more intuitive strategy is unexpected. One might <u>speculate</u> that during the low-R condition, the animated display did not possess the same degree of salience for the participants that it did in the med-R condition. Since two of the four cues had large dynamically changing reliability values in low-R as opposed to only a single cue (terrain) possessing the large reliability variance in med-R, the judges could not simply focus on the terrain cue. Instead, they had to distribute their attention over terrain and stealth cues in order to achieve more accurate judgments. During low-R, participants had to process significantly more information and this may have changed the manner in which judges were encoding the animation, from analyzing when only a single cue had the large reliability variance to intuitive processing when two cues had a large reliability variance.

4.4 Summary

The study found that changes in the reliability of the task environment were associated with decrements in performance of a multi-cue judgment task. The performance decrement during high and medium task reliability was primarily a function of the reduction in the participant's response consistency (cognitive control) in executing a learned judgment policy for integrating criterion information. During low task reliability, the performance decrement appeared in the form of reductions in the participant's knowledge of the task since a low diagnostic information source was wrongly weighted as highly informative of the criterion state.

Although it was clear that the reliability feed-forward information from the icon displays supported the judgment process as a main effect, the display format did not appear to matter in judgments produced during high-R conditions. Moreover, response time data provide some limited support to the notion that participants used analytical computation during high-R to render judgments. In contrast, format did seem to matter during the med-R condition with the graphic iconic format associated with superior judgment achievement scores (see figure 3). This finding was presumably attributable to this format successfully mapping the configural properties of the task through a spatial representation that participants were able to effectively understand and use. The reduced response time for the graphic display suggests that participants used an intuitive-anchored organizing principle during judgment. Finally, the animated icon display generated the greatest accuracy of the feed-forward display formats during low-R performance.

Clearly, a litany of important limitations exists in this study, which prevents any wholesale inferences to be drawn with regard to real judgment tasks employing iconic representation feed-forward applications. First, the present study used cues for the navigation task that were generated in a manner that produced very low cue inter-correlations. In addition, the information sources were represented as separate and distinct objects. This was done in order to simplify participant training on the judgment task and facilitate the evaluation of experimental manipulations that were aimed at the level of each cue reliability and magnitude elements used in judgments. Clearly, the pattern of results in much more ecological tasks, which employ correlated cues and multi-dimensional object formats, might be quite different.

Secondly, most inferences of cognitive mode in the present study tend to be circular. Although we stipulated that a rate measure of processing provides some independent assessment of an intuitive or analytical judgment state (i.e., organizing principle), this measure in itself is not nearly sufficient to define a mode independent that of the judgment indices themselves. As a result, only limited conclusions can be directed at particular cognitive modes during performance. Nevertheless, it was possible to reasonably differentiate modes of cognition based on data profiles, at least in part. For example, the possibility that participants resorted to a guessing strategy as opposed to executing an intuitive organizing principle was determined in relation to the matching (G) index values. A guessing strategy would not only generate poor knowledge control values (i.e., consistency) but poor task knowledge values as well (i.e., matching). In contrast, although an intuitive mode of cognition often suffers from lower control, matching values are often reasonably high.

Thirdly, it is presumed that information fidelity can be known and brought to bear on judgment problems within a variety of applied contexts. This assumption is equivalent to saying, in part, that valid procedures for obtaining real-time statistical assessments of reliability and/or uncertainty are available for use. Clearly, in many cases, this is not true for a number of reasons. A theoretical measurement problem for developing real-time probabilistic decision support systems lies in the manner in which information reliability and criterion reliability are modeled. Traditional measurement models cannot always address measurement error associated with the predictor variables or the notion of correlated errors that would be manifested in a real-time application of the display approach examined in this study (Lance, Baxter, & Mahan, in press). Reliability information generated from archival data, expert subjective estimates, or reports may partially fill this void until better procedures are developed for producing this information. Finally, recent work in the areas of virtual worlds and comprehensive simulations offer a method to study valid representations of complex decision environments that will support the detailed modeling of information properties (Elliott, Neville, Dalrymple, & Tower, 1997; Schiflett et al., 2004).

Future research might be directed at resolving some of the questions raised in this study. Representing information reliability as a specific property of an icon object display may help to create efficient and usable decision support devices of the kind described here. An additional research endeavor may include alternate representational schemes that use multi-modal approaches for displaying cue reliability information such as tactile and auditory information delivery, which, of course, will require significant changes in the methodology used in the present study. Finally, using measures that can independently validate the type of organizing principle being executed by decision makers will help develop iconic representations that induce appropriate and task-congruent cognitive processes in decision makers.

5. References

Anderson, J. R. Is Human Cognition Adaptive? *Behavioral and Brain Sciences* **1991**, *14*, 471-517.

Bailey, R. W. *Human Performance Engineering: A Guide for System Designers*, Englewood Cliffs, N.J.: Prentice Hall, 1982.

Bailey, R. W. *Human Performance Engineering: A Guide for System Designers.* Englewood Cliffs, N.J.: Prentice-Hall, 1989.

Balzer, W. K.; Doherty, M. E.; O'Connor, R. Effects of Cognitive Feedback on Performance. *Psychological Bulletin* **1989**, *106*, 410-433.

Brehmer, B. Inference Behavior in a Situation Where the Cues are Not Reliably Perceived. *Organizational Behavior and Human Performance* **1970**, *5* (4), 330-347.

Brehmer, B. Hypotheses about Relations Between Scaled Variables in the Learning of Probabilistic Inference Tasks. *Organizational Behavior and Human Performance* **1974**, *11* (1), 1-27.

Brehmer, B. Response Consistency in Probabilistic Inference Tasks. *Organizational Behavior and Human Performance* **1978**, *22* (1), 103-115.

Brehmer, B.; Joyce, C.R.B. *Human Judgment: The SJT View.* Amsterdam: North Holland, 1988.

Cannon-Bowers, J. A.; Salas, E.; Pruitt, J. S. Establishing the Boundaries of a Paradigm for Decision-Making Research. *Human Factors* **1996**, *38* (2), 193-205.

Carswell, C. M.; Wickens, C. D. Information Integration and the Object Display: An Integration of Task Demands and Display Superiority. *Ergonomics* **1987**, *30*, 511-527.

Cooksey, R. W. *Judgment Analysis: Theory, Methods, and Applications*. San Diego: Academic Press, 1996.

Coury, B. G.; Boulette , M. D.; Smith, R. A. Effect of Uncertainty and Diagnosticity on Classification of Multidimensional Data with Integral and Separable Displays of System Status. *Human Factors* **1989**, *31*, 551-569.

Dixon, S. R.; Wickens, C. D. *Reliability in Automated Aids for Unmanned Aerial Vehicle Flight Control: Evaluating a Model of Automation Dependence in High Workload.* (AHFD-04-05/MAAD-04-1). Savoy, IL: UIUC, AHFD, 2004a.

Dixon, S.; Wickens, C.D. Automation Reliability in Unmanned Aerial Vehicle Flight Control. *In Proceedings of the 5th Human Performance, Situation Awareness and Automation Technology Annual Meeting*, 2004b.

Doherty, M. E.; Sullivan, J. A. *Organizational Behavior and Human Decision Processes* **1989**, *43*, 136-144.

Dunwoody, P.; Marino, C.; Mahan, R. P.; Haarbauer, E. Cognitive Adaptation and its Consequences: A Test of Cognitive Continuum Theory. *Journal Behavioral Decision-Making* **2000**, *18,* 122-147.

Edwards, E.; Less, F. *The Human Operator in Process Control.* London: Taylor & Francis, 1974.

Elliott, L.; Neville, K.; Dalrymple, M.; Tower, S. Impact of Ambiguity on AWACS Team Communications and Decision Making. *Proceedings at the Ninth International Symposium on Aviation Psychology*, April, Columbus, OH, 1997.

Endsley, M. R.; Kiris, E. O. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors* **1995**, *37* (2), 381-394.

Garner, W. R. *The Processing of Information and Structure.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.

Garner, W. R. The Role of Configuration in the Identification of Visually Degraded Words. *Memory and Cognition* **1981**, *9* (5), 445-452.

Gentner, D.; Stevens, A. L. *Mental Models.* New Jersey: Lawrence Erlbaum Associates, 1983.

Hamm, R. M. Moment-by-Moment Variation in Experts' Analytic and Intuitive Cognitive Activity. *IEEE Transactions on Systems, Man, & Cybernetics* **1988**, *18* (5), 757-776.

Hammond, K. R. *The Integration of Research in Judgment and Decision Theory*; Report No. 226; Boulder, CO: Center for Research on Judgment and Policy, University of Colorado, 1980.

Hammond, K. R. *Principles of Organization in Intuitive and Analytical Cognition*; Report No. 231; Boulder, CO: Center for Research on Judgment and Policy, University of Colorado, 1981.

Hammond, K. R. Toward a Unified Approach to the Study of Expert Judgment. In J. Mumpower, L. Phillips, O. Renn, & R. Uppuluri (Eds.), *Expert judgment and expert systems* (pp. 1-16). Berlin: Springer-Verlag, 1987.

Hammond, K. R. Functionalism and Illusionism: Can Integration be Usefully Achieved? In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 231-269). Chicago, IL: University of Chicago Press, 1990.

Hammond, K. R. Naturalistic Decision Making from a Brunswikian Viewpoint: Its Past, Present and Future. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 205-227). Norwood, NJ: Ablex, 1993.

Hammond, K. R. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice.* New York: Oxford University Press, 1996.

Hammond, K. R.; Hamm, R. M.; Grassia, J.; Pearson, T. Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgments. *IEEE Transactions on Systems, Man and Cybernetics* **1987**, *SMC-17* (5), 753-770.

Hammond, K. R.; McClelland, G. H.; Mumpower, J. *Human Judgment and Decision Making: Theories, Methods and Procedures*. New York: Praeger, 1981.

Hammond, K. R.; Rohrbaugh, J.; Mumpower, J.; Adelman, L. Social Judgment Theory: Applications in Policy Formation. In M. Kaplan, & S. Schwartz (Eds.), *Human judgment and decision making in applied settings* (pp.23-66). New York: Academic Press, 1977.

Hammond, K. R.; Summers, D. A. Cognitive Control. *Psychological Review* **1972**, *79*, 58-67.

Hammond, K. R.; Wascoe, N. E. *Realizations of Brunswik's Representative Design.* San Francisco: Jossey-Bass Inc, 1980.

Hollenbeck, J. R.; Ilgen, D. R.; Sego, D. J.; Hedlund, J. Multilevel Theory of Team Decision Making: Decision Performance in Teams Incorporating Distributed Expertise. *Journal of Applied Psychology* **1995**, *80* (2), 292-316.

Kahneman, D.; Slovic, P.; Tversky, A. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.

Kahneman, D.; Tversky, A. On the Psychology of Prediction. *Psychological Review* **1973**, *80* (4), 237-251.

Kahneman, D.; Tversky, A. On the Interpretation of Intuitive Probability: A reply to Jonathan Cohen. *Cognition* **1979**, *7* (4), 409-411.

Kahneman, D.; Tversky, A. Variants of Uncertainty. *Cognition* **1982**, *11* (2), 143-157.

Kirschenbaum, S. S.; Arruda, J. E. Effects of Graphic and Verbal Probability Information on Command Decision Making. *Human Factors* **1994**, *36* (3), 406-418.

Klein, G. A.; Orasanu, J.; Calderwood, R.; Zsambok, C. *Decision Making in Action: Models and Methods.* Norwood, NJ: Ablex, 1993.

Kubovy, M.; Pomerantz, J. R. *Perceptual Organization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

Lance, C.; Baxter, D.; Mahan, R. Multi-Source Performance Measurement: A Reconceptualization. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges* (pp. xxx-xxx). Hillsdale, N J: Lawrence Erlbaum Associates, in press.

Mahan, R. P. The Effects of Extended Work on Multi-Dimensional Judgments of a Manufacturing Process. *International Journal of Human Factors in Manufacturing* **1991**, *1*, 155-165.

Mahan, R. P. Effects of Task Uncertainty and Continuous Performance on Knowledge Execution in Complex Decision Making. *International Journal of Computer Integrated Manufacturing* **1992**, *5*, 58-67.

Mahan, R. P. Stress Induced Strategy Shifts Toward Intuitive Cognition: A Cognitive Continuum Framework Approach. *Human Performance* **1994**, *7* (2), 85-118.

Marino, C. J.; Mahan, R. P. Supporting Nutrition-Related Decision-Making Through Proximity Compatibility Display Principles. *Human Factors* (pp. xxx-xxx), in press.

Massimino, M. J.; Sheridan, T. B. Teleoperator Performance with Varying Force and Concealment Feedback. *Human Factors* **1994**, *36* (1) 145-157.

Moray, N. Monitoring Behavior and Supervisory Control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 178-194). New York: Wiley, 1986.

Orasanu, J.; Salas, E. Team Decision Making in Complex Environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3-20). Norwood, NJ: Ablex, 1993.

Parasuraman, R.; Molloy, R.; Singh, I. Performance Consequences of Automation-Induced "Complacency". *International Journal of Aviation Psychology* **1993**, *3* (1), 1-23.

Pomerantz, J. R. Perceptual Organization in Information Processing. In K. M. & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141-180). Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

Pomerantz, J. R.; Pristach, E. A. Emergent Features, Attention, and Perceptual Glue in Concealment Form Perception. *Journal of Experimental Psychology: Human Perception and Performance* **1989**, *15* (4), 635-649.

Rasmussen, J. A Cognitive Engineering Approach to the Modeling of Decision Making and its Organization in Process Control, Emergency Management, CAD/CAM, Office Systems, and

Library Systems. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (pp. 223-257). Greenwich, CT: JAI Press, 1988.

Rouse, W. *Engineering Models of Human Performance.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.

Schiflett, S.; Elliott, L.R.; Salas, E.; Coovert, M. *Scaled Worlds: Development and Utilization.* Ashgate Publishing: London, 2004.

Searcy, S. *The Effects of Continuous Performance and Task Uncertainty on Risky Multicue Probability Judgments.* Master's Thesis, University of Georgia, Athens, GA, 1994.

Sheridan, T. B. Toward a General Model of Supervisory Control. In T. B. Sheridan, & G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 271-282). New York: Plenum, 1976.

Sheridan, T. B. *Telerobotics, Automation, and Human Supervisory Control.* Cambridge, MA: MIT Press, 1992.

Simon, H. A. Rationality as a Process and Product of Thought. *American Economic Review* **1978**, *68* (1-16).

Simon, H. A. Alternative Visions of Rationality. In Moser, P. K. (Ed.), *Rationality in action: Contemporary approaches* (pp. 55-82). Cambridge, MA: Cambridge University Press, 1990.

Slovic, P.; Fischoff, B.; Lichtenstein, S. Behavioral Decision Theory. *Annual Review of Psychology* **1977**, *28*, 1-39.

Stewart, T. R. Improving Reliability of Judgmental Forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81-106). New York: Kluwer Academic Publishers, 2001.

Stewart, T. R.; Lusk, C. M. Seven Components of Judgmental Forecasting Skill: Implications for Research and Improvement of Forecasts. *Journal of Forecasting* **1994**, *13*, 579-599.

Tucker, L. R. A Suggested Alternative Formulation in the Developments by Hursch, Hammond, & Hursch and by Hammond, Hursch, & Todd. *Psychological Review* **1964**, *71*, 528-530.

Tversky, A.; Kahneman, D. Judgment Under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124-1131.

Tversky, A.; Kahneman, D. Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* **1983**, *90* (4) 293-315.

U.S. Army Training and Doctrine Command. *Army Readiness Training* (TRADOC Pam 350-4, Coordinating Draft). Fort Monroe, VA, 1989.

NO. OF COPIES	ORGANIZATION
1 (PDF ONLY)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	US ARMY RSRCH DEV & ENGRG CMD SYSTEMS OF SYSTEMS INTEGRATION AMSRD SS T 6000 6TH ST STE 100 FORT BELVOIR VA 22060-5608
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC IMS 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
2	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS OK T 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR M DR M STRUB 6359 WALKER LANE SUITE 100 ALEXANDRIA VA 22310
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MA J MARTIN MYER CENTER RM 2D311 FT MONMOUTH NJ 07703-5630
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MC A DAVISON 320 MANSCEN LOOP STE 166 FT LEONARD WOOD MO 65473-8929
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MD T COOK BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290

NO. OF COPIES	ORGANIZATION
1	COMMANDANT USAADASCH ATTN ATSA CD ATTN AMSRD ARL HR ME MS A MARES 5800 CARTER RD FT BLISS TX 79916-3802
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MO J MINNINGER BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MM DR V RICE BLDG 4011 RM 217 1750 GREELEY RD FT SAM HOUSTON TX 78234-5094
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MG R SPINE BUILDING 333 PICATINNY ARSENAL NJ 07806-5000
1	ARL HRED ARMC FLD ELMT ATTN AMSRD ARL HR MH C BURNS BLDG 1467B ROOM 336 THIRD AVENUE FT KNOX KY 40121
1	ARMY RSCH LABORATORY - HRED AVNC FIELD ELEMENT ATTN AMSRD ARL HR MJ D DURBIN BLDG 4506 (DCD) RM 107 FT RUCKER AL 36362-5000
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MK MR J REINHART 10125 KINGMAN RD FT BELVOIR VA 22060-5828
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MV HQ USAOTC S MIDDLEBROOKS 91012 STATION AVE ROOM 111 FT HOOD TX 76544-5073
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MY M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FT HUACHUCA AZ 85613-7069
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MP D UNGVARSKY BATTLE CMD BATTLE LAB 415 SHERMAN AVE UNIT 3 FT LEAVENWORTH KS 66027-2326

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MJK J HANSBERGER JFCOM JOINT EXPERIMENTATION J9 JOINT FUTURES LAB 115 LAKEVIEW PARKWAY SUITE B SUFFOLK VA 23435
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MQ M R FLETCHER US ARMY SBCCOM NATICK SOLDIER CTR AMSRD NSC SS E BLDG 3 RM 341 NATICK MA 01760-5020
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MY DR J CHEN 12423 RESEARCH PARKWAY ORLANDO FL 32826
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MS MR C MANASCO SIGNAL TOWERS RM 303A FORT GORDON GA 30905-5233
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MU M SINGAPORE 6501 E 11 MILE RD MAIL STOP 284 BLDG 200A 2ND FL RM 2104 WARREN MI 48397-5000
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MF MR C HERNANDEZ BLDG 3040 RM 220 FORT SILL OK 73503-5600
10	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MW E REDDEN BLDG 4 ROOM 332 FT BENNING GA 31905-5400
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MN R SPENCER DCSFDI HF HQ USASOC BLDG E2929 FORT BRAGG NC 28310-5000
1	ARMY G1 ATTN DAPE MR B KNAPP 300 ARMY PENTAGON ROOM 2C489 WASHINGTON DC 20310-0300

NO. OF COPIES	ORGANIZATION
	ABERDEEN PROVING GROUND
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK (TECH LIB) BLDG 4600
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK TP S FOPPIANO BLDG 459
1	ARL HRED ATTN AMSRD ARL HR MB F PARAGALLO BLDG 459

Vicente, K. J.; Rasmussen, J. Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, & Cybernetics* **1992**, *22* (4), 589-606.

vonWinterfeldt, D.; Edwards, W. *Decision Analysis in Behavioral Research.* Cambridge, England: Cambridge University Press, 1986.

Wickens, C. D. *Engineering Psychology and Human Performance.* New York: Harper Collins, 1992.

Wickens, C. D.; Andre, A. D. Proximity Compatibility and Information Displays: Effects of Color, Space, and Objectness in Information Integration. *Human Factors* **1990**, *32*, 61-77.

Wickens, C. D.; Carswell, C. M. The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design. *Human Factors* **1995**, *37* (3), 473-494.

Wickens, C. D.; Gempler, K.; Morphew, M. Workload and Reliability of Predictor Displays in Aircraft Traffic Avoidance. *Transportation Human Factors* **2000**, *2*, 99-126.

Wiener, E. L. Cockpit Automation. In E. L. Wiener, & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp 67-89). San Diego: Academic Press, 1988.

Wiener, N. *Cybernetics*. Cambridge, MA: MIT Press, 1948.

Williams, C.; Mahan R. P. (in press). Teams Leading Teams: An Experimental Investigation of Leadership in Multi-Team Systems. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), *Performance Measurement: Current Perspectives and Future Challenges* (pp. xxx-xxx). Hillsdale, NJ: Lawrence Erlbaum Associates.

Woods, D. Coping With Complexity: The Psychology of Human Behavior in Complex Systems. In H. B.Goodstein, and S. E. Olsen (Eds.), *Tasks, errors and mental models* (pp. 13-33). Philadelphia, PA: Taylor and Francis, 1988.

York, K. M.; Doherty, M. E.; Kamouri, J. The Influence of Cue Unreliability on a Judgment in a Multiple-Cue Probability Learning Task. *Organizational Behavior and Human Decision Processes* **1987**, *39*, 303-317.

Zsambok, C.; Klein, G. *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.